

**PCT**WELTORGANISATION FÜR GEISTIGES EIGENTUM  
Internationales BüroINTERNATIONALE ANMELDUNG VERÖFFENTLICHT NACH DEM VERTRAG ÜBER DIE  
INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT)(51) Internationale Patentklassifikation <sup>6</sup> :  
G10L 3/00

A1

(11) Internationale Veröffentlichungsnummer: WO 98/10413

(43) Internationales  
Veröffentlichungsdatum: 12. März 1998 (12.03.98)

(21) Internationales Aktenzeichen: PCT/DE97/01920

(22) Internationales Anmeldedatum: 2. September 1997 (02.09.97)

(30) Prioritätsdaten:  
196 35 754.3 3. September 1996 (03.09.96) DE(71) Anmelder (für alle Bestimmungsstaaten ausser US): SIEMENS  
AKTIENGESELLSCHAFT [DE/DE]; Wittelsbacher Platz 2,  
D-80333 München (DE).

(72) Erfinder; und

(75) Erfinder/Anmelder (nur für US): NIEDERMAIR, Gerhard  
[DE/DE]; Ludwig-Thoma-Strasse 9, D-85256 Vierkirchen  
(DE). KARSTEN, Peter [DE/DE]; Andreas-Sammer-Strasse  
2, D-82061 Neuried (DE). HARENGEL, Steffen [DE/DE];  
Mitterfeldstrasse 59b, D-80689 München (DE). MURR,  
Florian [DE/DE]; Görresstrasse 11, D-80798 München  
(DE).(81) Bestimmungsstaaten: CN, US, europäisches Patent (AT, BE,  
CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL,  
PT, SE).

Veröffentlicht

*Mit internationalem Recherchenbericht.  
Vor Ablauf der für Änderungen der Ansprüche zugelassenen  
Frist. Veröffentlichung wird wiederholt falls Änderungen  
eintreffen.*

(54) Title: SPEECH-PROCESSING SYSTEM AND METHOD

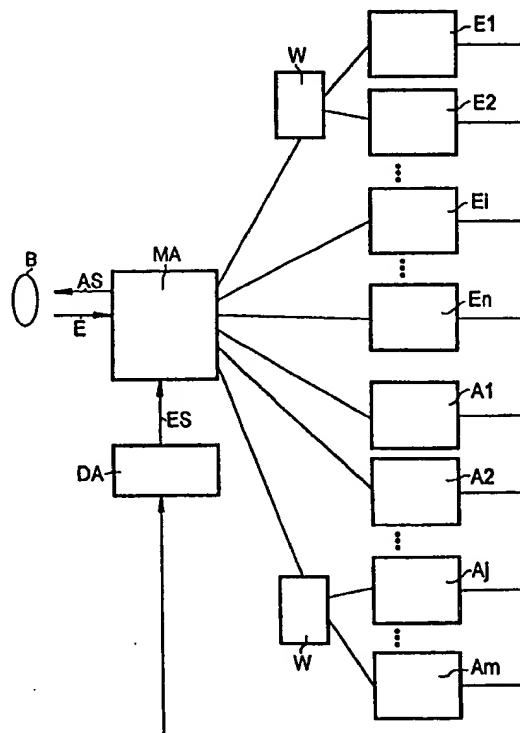
(54) Bezeichnung: SPRACHVERARBEITUNGSSYSTEM UND VERFAHREN ZUR SPRACHVERARBEITUNG

(57) Abstract

The invention concerns a speech-processing system which comprises an optional number of speech-recognition modules (E<sub>i</sub>, i = 1..n) and speech-output modules (A<sub>j</sub>, j = 1..m). Depending on the application, the modules, which are each provided for a given type of speech recognition or speech output, are selected, activated and parametrized by a module-selection device (MA) as a function of an input signal (ES), such that the selected modules are used according to their respective configuration for speech-recognition or speech-output purposes.

(57) Zusammenfassung

Ein Sprachverarbeitungssystem weist eine beliebige Anzahl von Spracherkennungsmodulen (E<sub>i</sub>, i = 1..n) und Sprachausgabemodulen (A<sub>j</sub>, j = 1..m) auf. Die jeweils für eine bestimmte Art der Spracherkennung bzw. Sprachausgabe vorgesehenen Module werden je nach Anwendungssituation abhängig von einem Eingangssignal (ES) von einem Mittel (MA) zur Auswahl der Module ausgewählt und aktiviert und parametrisiert, so daß die ausgewählten Module entsprechend ihrer Ausgestaltung zur Spracherkennung bzw. Sprachausgabe verwendet werden.



### LEDIGLICH ZUR INFORMATION

Codes zur Identifizierung von PCT-Vertragsstaaten auf den Kopfbögen der Schriften, die internationale Anmeldungen gemäss dem PCT veröffentlichen.

AL	Albanien	ES	Spanien	LS	Lesotho	SI	Slowenien
AM	Armenien	FI	Finnland	LT	Litauen	SK	Slowakei
AT	Österreich	FR	Frankreich	LU	Luxemburg	SN	Senegal
AU	Australien	GA	Gabun	LV	Lettland	SZ	Swasiland
AZ	Aserbaidshan	GB	Vereinigtes Königreich	MC	Monaco	TD	Tschad
BA	Bosnien-Herzegowina	GE	Georgien	MD	Republik Moldau	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagaskar	TJ	Tadschikistan
BE	Belgien	GN	Guinea	MK	Die ehemalige jugoslawische Republik Mazedonien	TM	Turkmenistan
BF	Burkina Faso	GR	Griechenland	ML	Mali	TR	Türkei
BG	Bulgarien	HU	Ungarn	MN	Mongolei	TT	Trinidad und Tobago
BJ	Benin	IE	Irland	MR	Mauretanien	UA	Ukraine
BR	Brasilien	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Island	MX	Mexiko	US	Vereinigte Staaten von Amerika
CA	Kanada	IT	Italien	NE	Niger	UZ	Usbekistan
CF	Zentralafrikanische Republik	JP	Japan	NL	Niederlande	VN	Vietnam
CG	Kongo	KE	Kenia	NO	Norwegen	YU	Jugoslawien
CH	Schweiz	KG	Kirgisistan	NZ	Neuseeland	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Demokratische Volksrepublik Korea	PL	Polen		
CM	Kamerun	KR	Republik Korea	PT	Portugal		
CN	China	KZ	Kasachstan	RO	Rumänien		
CU	Kuba	LC	St. Lucia	RU	Russische Föderation		
CZ	Tschechische Republik	LI	Liechtenstein	SD	Sudan		
DE	Deutschland	LK	Sri Lanka	SE	Schweden		
DK	Dänemark	LR	Liberia	SG	Singapur		
EE	Estland						

## Beschreibung

5

Sprachverarbeitungssystem und Verfahren zur Sprachverarbeitung

10

In üblichen Sprachverarbeitungssystemen beispielsweise für telefonbasierte Anwendungen werden verschiedenste Möglichkeiten der sprachlichen Eingabe und Ausgabe angeboten. Dies reicht von der Erkennung von einzeln gesprochenen Ziffern und einer meist begrenzten Anzahl von Kommandowörtern bis zur Eingabemöglichkeit fließend gesprochener Äußerungen. Die Ausgabemöglichkeiten reichen vom Ausgeben komplett aufgenommener Äußerungen bis zu rein synthetisch erzeugter akustischer Ausgabe von Texten aus beispielsweise abgespeicherten Phonemen.

15

Anwendungen von Sprachverarbeitungssystemen im industriellen, öffentlichen sowie im privaten Bereich sind mit sehr unterschiedlichen Anforderungen sowohl an die Spracheingabe als auch an die Sprachausgabe konfrontiert. Dies ist bedingt durch die Vielfalt der Anwendungsmöglichkeiten und der damit von den Sprachverarbeitungssystemen zu lösenden sprachlichen Aufgaben.

25

Dies gilt ebenso für Sprachdialogsysteme, bei denen das Sprachverarbeitungssystem mit dem Benutzer des Sprachverarbeitungssystems in einem beliebig ausgestalteten Dialog steht.

30

Bei einer Anwendung im Bereich von Sprachverarbeitungssystemen kann es beispielsweise erforderlich sein, daß zu einem Zeitpunkt vom Benutzer eine Ziffer einzeln abgefragt wird, zu einem anderen Zeitpunkt innerhalb desselben Dialoges aber von dem Benutzer ein Datum eingegeben werden kann und vom Sprachverarbeitungssystem verarbeitet werden soll.

35

Ebenso kann bei der Ausgabe an einer Stelle die Ausgabe einer Sprachkonserve ausreichend sein, während an anderer Stelle eines Anwendungsszenarios durch die hohe Variabilität des auszugebenden Textes eine Sprachsynthese geeigneter ist.

Desweiteren können zu verschiedenen Zeitpunkten verschiedene Leistungsmerkmale von Spracherkennern vom Betreiber als erforderlich erachtet werden, wie z. B. die Unterbrechbarkeit der Systemausgabe wünschenswert sein kann, wohingegen an anderer Stelle jedoch der Benutzer die Ausgabe vollständig anhören sollte.

Um eine möglichst effiziente Verwendung der spezialisierten, bekannten Algorithmen zur Spracherkennung und zur Sprachausgabe und auch Sprachdialogsysteme mit komplexeren Erkennungseigenschaften und Ausgabeeigenschaften als beispielsweise zur einfachen Ziffernerkennung oder nur zur isolierten Erkennung von einer kleinen Anzahl von Kommandowörtern, war es bisher üblich, die Systeme in einer Weise zu erstellen, daß durch die besondere Auswahl von technischen Methoden und spezifischen softwaretechnischen Maßnahmen den spezifischen Anforderungen an die Spracherkennung bzw. an die Sprachausgabe Rechnung getragen wurde.

Dies erfordert jedoch einen großen personellen Aufwand von Experten, die für jede neue Anwendungssituation in der Regel neu zu erbringen ist. Durch diese Vorgehensweise wird eine effiziente wirtschaftliche Nutzung von Sprachverarbeitungssystemen erheblich behindert.

Im Telefonbereich existieren Dialogsysteme, bei denen eine Auswahl beispielsweise über sog. DTMF-Töne und einfache Nennung von Ziffern oder einiger, weniger vorgegebener Kommandoworte erfolgt.

Jedoch ist der Aufwand für eine neue Entwicklung eines Sprachverarbeitungssystems bzw. eines Sprachdialogsystems mit den bisherigen Verfahren und Systemen sehr aufwendig.

- 5 Aus [3] ist ein Verfahren zur Verbesserung der Spracherken-  
nung unter Berücksichtigung verschiedener digitaler Codierbe-  
dingungen bekannt. Dabei wird für eine Mehrzahl von Spracher-  
kennungsmodulen, die alle den gleichen Spracherkennungsalgo-  
rithmus verwenden, jedoch möglichst optimal adaptiert an un-  
10 terschiedliche Kanalcodierungsverfahren für empfangene Si-  
gnale das jeweils verwendete Kanalcodierungsverfahren erkannt  
und das jeweilig entsprechende Spracherkennungsmodul ausge-  
wählt.
- 15 Aus [6] ist ein Verfahren und eine Anordnung zur Generierung  
einer Vielzahl elektrischer Signale bekannt. Bei dieser An-  
ordnung werden eine Vielzahl von Sprachausgabeeinheiten glei-  
chen Aufbaus miteinander kombiniert zu der Gesamtanordnung.
- 20 Somit liegt der Erfindung das Problem zugrunde, ein Sprach-  
verarbeitungssystem sowie ein Verfahren zur Sprachverarbei-  
tung anzugeben, mit dem eine optimale, kontextangepaßte Nut-  
zung von Spracherkennungsmodulen und/oder Sprachausgabemodu-  
len unter Verminderung benötigter Rechenkapazität erreicht  
25 wird.

Das Problem wird durch das Sprachverarbeitungssystem gemäß  
Patentanspruch 1 sowie durch das Verfahren gemäß Patentan-  
spruch 8 gelöst.

30

- Das Sprachverarbeitungssystem weist mehrere Spracherkennungs-  
module und/oder Sprachausgabemodule auf, die jeweils für eine  
vorgebbare Art einer Spracherkennung bzw. einer Sprachausgabe  
vorgesehen sind. Ferner weist das Sprachverarbeitungssystem  
35 ein Mittel zur Auswahl eines Spracherkennungsmoduls und/oder  
Sprachausgabemoduls auf. Durch das Mittel wird abhängig von  
einem Eingangssignal, welches dem Mittel zugeführt wird, das

jeweilige Spracherkennungsmodul bzw. Sprachausgabemodul ausgewählt, aktiviert und das entsprechende Sprachsignal bzw. die erforderlichen Parameter werden dem jeweiligen Modul zugeführt. Die Spracherkennung bzw. die Sprachausgabe wird dann  
5 von dem jeweiligen Modul ausgeführt.

Bei dem Verfahren wird ein Eingangssignal analysiert mit dem beschrieben wird, welche Art von Spracherkennung bzw. Sprachausgabe benötigt wird. Abhängig von dem Eingangssignal wird  
10 mindestens ein Spracherkennungsmodul und/oder mindestens ein Sprachausgabemodul für die jeweils durch das Eingangssignal beschriebene Spracherkennung bzw. Sprachausgabe, aktiviert. Mit dem jeweils ausgewählten bzw. aktivierten Modul wird die jeweilige Art der Spracherkennung bzw. Sprachausgabe durchge-  
15 führt.

Sowohl das Sprachverarbeitungssystem als auch das Verfahren weisen eine Vielzahl erheblicher Vorteile gegenüber dem Stand der Technik auf.

20

Der Bedarf an Rechenkapazität bzw. an Verarbeitungsressourcen für die Spracherkennung bzw. zur Sprachausgabe wird erheblich reduziert, da abhängig von der jeweils benötigten Anwendungssituation das für die Situation optimal angepaßte Modul zur  
25 Spracherkennung bzw. zur Sprachausgabe verwendet wird.

Auch bei der Erstellung von Sprachdialogsystem ist das Sprachverarbeitungssystem zur Entwicklungsunterstützung sehr vorteilhaft einsetzbar. Besondere Vorteile bietet dieses  
30 Sprachverarbeitungssystem in diesem Bereich, da eine schnelle und einfache Anpassung der vorhandenen technischen Spracherkennungsverfahren und Sprachausgabeverfahren an die jeweiligen spezifischen Erfordernisse der Anwendungssituation möglich ist.

35

Weiterhin wird eine flexible, der spezifischen Dialogssituation angepaßte Nutzung von deren unterschiedlichen Fähigkei-

ten und Leistungsmerkmalen erreicht. Auch ist eine flexible Anpassung der Spracherkennungsleistung und Sprachausgabeleistung an die bei den Benutzern in der spezifischen Dialogsituation verfügbaren Ressourcen, beispielsweise der verfügbaren Rechnerkapazität möglich.

Somit stellt das Sprachverarbeitungssystem ein bedeutendes Mittel dar zur bedarfsgerechten Konfiguration und Anpassung von Spracheingabe und Sprachausgabe in verschiedensten Anwendungssituationen.

Vorteilhafte Weiterbildungen der Erfindung ergeben sich aus den abhängigen Ansprüchen.

Es ist in einer Weiterbildung des Sprachverarbeitungssystems sehr vorteilhaft, bestimmte Algorithmen, also bestimmte Teile der Module gemeinsam zu verwenden, wodurch eine redundante Realisierung lediglich eigentlich nur einmal benötigter Algorithmen vermieden wird.

Auch ist eine vorteilhafte Weiterbildung darin zu sehen, daß mehrere Spracherkennungsmodule gemeinsame Ressourcen verwenden, beispielsweise gemeinsame Lexika für die Spracherkennung. Dies führt zu einer erheblichen Einsparung von Ressourcen, z. B. von Speicherbedarf.

Weiterhin ist es vorteilhaft, eine Dialoglaufsteuerung vorzusehen, mit der ein Dialog des Sprachverarbeitungssystems mit einem Benutzer realisiert werden kann.

Ein Ausführungsbeispiel der Erfindung ist in den Figuren dargestellt und wird im weiteren näher erläutert.

Es zeigen

Fig. 1 das Sprachverarbeitungssystem mit einigen Weiterbildungen;

Fig. 2 ein Blockdiagramm, in dem die Dialogablauf-

- steuerung dargestellt ist;
- Fig. 3 ein Ablaufdiagramm, in dem ein mögliches Verfahren, welches zur Dialogablaufsteuerung durchgeführt wird, dargestellt ist;
- 5 Fig. 4 ein Ablaufdiagramm, in dem das erfindungsgemäße Verfahren dargestellt ist.

In Fig. 1 ist das Sprachverarbeitungssystem dargestellt. Das Sprachverarbeitungssystem weist mindestens ein Spracherkennungsmodul  $E_1, E_2, E_i, \dots, E_n$  auf. Mit einem Index  $i$ , einer beliebigen Zahl zwischen 1 und  $n$ , wird jeweils das Spracherkennungsmodul  $E_i$  eindeutig gekennzeichnet. Mit  $n$  wird eine Anzahl von Spracherkennungsmodulen in dem Sprachverarbeitungssystem bezeichnet.

15

Ferner weist das Spracherkennungssystem zusätzlich oder alternativ mindestens ein Sprachausgabemodul  $A_1, A_2, A_j, \dots, A_m$  auf. Mit einem zweiten Index  $j$ , einer beliebigen Zahl zwischen 1 und  $m$ , wird jedes Sprachausgabemodul  $A_j$  eindeutig gekennzeichnet. Mit  $m$  wird eine Anzahl von Sprachausgabemodulen in dem Sprachverarbeitungssystem bezeichnet.

20

Ferner weist das Sprachverarbeitungssystem ein Mittel MA zur Auswahl der Spracherkennungsmodule  $E_i$  bzw. der Sprachausgabemodule  $A_j$  auf. Das Mittel MA zur Auswahl der Module ist mit den einzelnen Modulen  $E_i, A_j$  gekoppelt, so daß von dem Mittel MA zur Auswahl der Module  $E_i, A_j$  an die einzelnen Module  $E_i, A_j$  Signale gesendet werden können, beispielsweise zur Aktivierung der einzelnen Module  $E_i, A_j$  oder auch zur Übergabe von Parametern, die von den einzelnen Modulen zur Durchführung der Spracherkennung oder der Sprachausgabe benötigt werden.

30

#### Spracherkennungsmodule $E_i$

35

Grundlagen über die Spracherkennung sind beispielsweise in den Dokumenten [1] und [2] beschrieben.



Die einzelnen Spracherkennungsmodule  $E_1, E_2, \dots, E_n$  sind Module, die jeweils auf eine bestimmte Art der Spracherkennung spezialisiert sind.

5

Die einzelnen Spracherkennungsmodule  $E_i$  können beispielsweise durch Hardware oder auch durch einzelne Softwareprozeduren realisiert sein.

10 Im folgenden wird eine keineswegs abschließend zu betrachtende Übersicht über vorgesehene Spracherkennungsmodule  $E_i$  gegeben, die im Rahmen des Sprachverarbeitungssystems verwendet werden können.

15 - Einzelziffernerkenner

Ist ein Spracherkennungsmodul  $E_i$  als ein Einzelziffernerkenner ausgebildet, so ist es derart ausgestaltet, als Eingabe ein einzelnes, vom Benutzer B des Sprachverarbeitungssystems eingesprochenes Sprachsignal, welches eine einzige Ziffer repräsentiert, aus einer Auswahl von Ziffern, zu verarbeiten. Dabei unterscheiden sich die Einzelziffernerkenner üblicherweise in der Modellierung der akustischen Muster. Es sind Systeme bekannt, in denen von verschiedenen Sprechern die akustischen Muster für ein ganzes Wort gespeichert werden und mittels Vergleichsverfahren wie dem sog. Dynamic Programming oder dem Dynamic Time Warping oder auch unter Verwendung neuronaler Netze beim Erkennungsvorgang das am besten passende Wort ermittelt wird. Andere bekannte Module zur Einzelziffernerkennung basieren auf dem Prinzip der Hidden-Markov-Modelle. Bei diesen Prinzip der Einzelziffernerkennung wird die Lautfolge der von mehreren Sprechern gesprochenen Wörter als Zustandsfolge mit Variable und vorgegebener Länge abgespeichert. Bei der Erkennung der Wörter, in diesem Fall der Ziffer, wird der geringste Abstand zu der abgespeicherten Wortfolge ermittelt. Bei Varianten dieses Prinzips werden in Hidden-Markov-Modellen die Folgen der einzelnen phonetischen Einheiten abgespeichert, die aus den Daten mehrerer Sprecher

gewonnen werden, in der Regel Ziffern. Zur Verminderung benötigter Rechenkapazität kann bei einem Einzelziffernerkennungsausgenutzt werden, daß auf keine der gespeicherten lautlichen Folgen eine weitere Folge auftreten kann. Aus diesem Grund  
5 ist in diesem Fall keine rechenaufwendige Detektion von Wortgrenzen erforderlich. Auch ist der zu untersuchende Wortschatz relativ gering, da lediglich die Ziffern untersucht werden müssen.

10 - Ziffernkettenerkennung

Ist ein Spracherkennungsmodul Ei als Ziffernkettenerkennungausgestaltet, so ist das Spracherkennungsmodul in der Regel mit Erkennern für kontinuierliche Sprache gleichzusetzen. Eine Anzahl von in einem digitalen Lexikon definierten, vorgegebenen Wörtern, in diesem Spezialfall Ziffern, kann in beliebiger Reihenfolge hintereinander von einem Benutzer B des Sprachverarbeitungssystems geäußert werden. Für diesen Fall  
15 ist eine Detektion der Grenzen zwischen den einzelnen Worten erforderlich. Dies kann entweder durch ein Verfahren zur expliziten Detektion der Wortgrenzen erfolgen, oder die Wortgrenzen werden von dem Algorithmus der Wortsuche festgelegt. Bei der Wortsuche werden parallel alle Ziffern im Eingabesignal untersucht und verschiedene Enden hypothetisiert. An jedem Ende kann im Signal eine weitere Ziffer gesucht werden.  
20 Dazu wird ein Suchraum aufgebaut, der die verschiedenen Alternativen gleichzeitig verfolgt und sie nach vorgebbaren Qualitätskriterien ordnet. In diesem Zusammenhang kann beispielsweise das Prinzip der Dynamischen Programmierung verwendet werden, wie es in dem Dokument [1] beschrieben ist.  
25 Ein solches Kriterium kann der Abstand der beobachteten Teile des Signals zu den Referenzmustern, den einzelnen gespeicherten Ziffern sein. Ziffernkettenerkennung zeichnen sich in der Regel von kontinuierlichen Erkennern in der besonderen Art der Modellierung der akustischen Referenzmuster aus, bei denen bei der Ziffer eine spezifische Artikulationsweise oder  
35 typische Längen der einzelnen Ziffern ausgenutzt werden. Durch eine Modellierung von Sprachpausen oder durch ein Ent-

scheidungskriterium, z. B. die im Signal sichtbare Energie, kann in unterschiedlichen Heuristiken das Ende einer solchen Folge detektiert werden.

- 5 - Erkennung von Wörtern aus einem begrenzten Vokabular  
Ist ein Spracherkennungsmodul Ei als ein Spracherkenner für bestimmte häufige Kommandoworte aus einem begrenzten Vokabular, dem sog. Kernvokabular ausgestaltet, so werden algorithmisch die entsprechenden Verfahren angewendet wie auch zur  
10 Einzelziffernerkennung, jedoch mit dem Unterschied, daß das Vokabular um die entsprechenden Worte erweitert ist bzw. daß sich das Vokabular, mit den entsprechenden gespeicherten Repräsentanten der einzelnen Worte von den Ziffern unterscheidet. Das speziell erweiterte bzw. ersetzte Vokabular ist für  
15 den Spracherkenner vorgegeben und in der Regel mit einer größeren Anzahl von Sprechern trainiert.

- Erkennung eines Einzelwortes aus einem unbegrenzten Vokabular  
20 Zur Erkennung fließend gesprochener Sprache mit einem unbegrenzten Vokabular ist ein Spracherkennungsmodul Ei derart ausgestaltet, daß die Worte des Vokabulars dem Spracherkennungsalgorithmus, der in dem Spracherkennungsmodul Ei realisiert wird, nicht bekannt sind, sondern von einem Benutzer B  
25 zuvor bestimmt werden. Der Algorithmus zur Spracherkennung basiert üblicherweise auf der Erkennung einer Folge von lautlichen Einheiten, den sog. Phonemen oder anderen Wortuntereinheiten, aus denen sich die Worte des zu erkennenden Vokabulars zusammensetzen lassen. In dem Lexikon des Spracherkennungsmoduls Ei werden in einem Speicher die das jeweilige  
30 Wort kennzeichnenden lautlichen Folgen abgespeichert. Die Laute können in ihrer Modellierung auch noch weitere Unterseinheiten, wie z. B. Beginn des Phonems, Mitte des Phonems, Ende des Phonems, Übergang des Phonems, etc. aufweisen. Das  
35 Spracherkennungsverfahren sucht aus dem beobachteten Sprachsignal wiederum die Folge mit dem geringsten Abstand zu den Referenzdaten auf. Der Einzelworterkenner kann algorithmisch

den Umstand ausnutzen, daß auf keine der im Inventar abgelegten lautlichen Folgen eine weitere Folge auftreten kann.

- Erkennung fließend gesprochener Sprache mit einem unbegrenzten Vokabular

5 Ist ein Spracherkennungsmodul Ei zur Erkennung fließend gesprochener Äußerungen des Benutzers mit einem unbegrenzten Vokabular ausgestaltet, so unterscheidet sich das Spracherkennungsmodul Ei von einem Spracherkennungsmodul Ei zur Ziffernkettenerkennung in der Regel in der Art der Modellierung der lautlichen Einheiten. Ist das lexikalische Inventar, also das Vokabular, unbegrenzt, was auch als ein offenes lexikalisches Inventar bezeichnet wird, erfolgt die akustische Modellierung meist auf lautlichen Einheiten, den Phonemen, oder  
10 auf Wortuntereinheiten, aus denen sich das zu erkennende Vokabular zusammensetzen läßt. Ebenso wie bei dem Spracherkennungsmodul Ei für kontinuierliche Ziffern werden bei dieser Ausgestaltung des Spracherkennungsmoduls die Grenzen zwischen den Wörtern durch eine geeignete Suche bestimmt. Dies erfolgt  
15 beispielsweise durch eine Modellierung von Sprachpausen oder durch ein Entscheidungskriterium, z. B. die im Signal sichtbare Energie. Das Spracherkennungsmodul kann die wahrscheinlichste, oder mit Erweiterungen im Suchalgorithmus eine vorgebbare Anzahl wahrscheinlichster Folgen von Wörtern ausgeben. Die Suche bei kontinuierlichen Spracherkennungsmodulen  
20 wird meistens durch eine Modellierung der für einen bestimmten Sprachausschnitt typischen oder möglichen Wortfolgen, den sog. Language-Modellen unterstützt. Unter Language-Modellen sind beispielsweise grammatische Modellierungen der Wortfolgen oder Wortlisten oder statistische Modellierungen im Sinne  
25 einer sog. N-Gram-Analyse zu verstehen.

- Erkennung von vorgegebenen Wortkombinationen

Zur Erkennung von vorgegebenen Wortkombinationen ist ein  
35 Spracherkennungsmodul Ei derart ausgestaltet, daß beispielsweise Datumsangaben, Uhrzeitangaben, Maßangaben, Zahlenangaben, etc. als spezielles Vokabular verwendet werden. Bei die-

sen Spracherkennungsmodulen Ei handelt es sich üblicherweise um Spracherkennungsmodule Ei für kontinuierliche Sprache, deren lexikalisches Inventar aber auf die vorgegebenen Wortkombinationen aus den jeweiligen Bereichen, wie beispielsweise  
5 eine Datumsangabe oder eine Maßangabe zugeschnitten ist. Des weiteren kann die Wortfolgensuche durch eine Modellierung der möglichen Wortfolgen wie bei einem kontinuierlichen Spracherkennungsmodul Ei unterstützt werden, wobei sich diese Modellierung dadurch auszeichnet, daß sie im wesentlichen die zu  
10 erwartenden Ausdrucksweisen in einem dieser sprachlichen Untergebiete, z. B. Datumsangabe, besonders berücksichtigt.

- Schlüsselwörterkennung

Bei einem Spracherkennungsmodul Ei zur Schlüsselwörterkennung  
15 können in einem Spracheingabesignal ein oder mehrere vorgegebene Worte, die in einem Lexikon zuvor gespeichert wurden, detektiert werden. Die verschiedenen Möglichkeiten zur Ausgestaltung der Spracherkennungsmodule zur Schlüsselwörterkennung sind beispielsweise in dem Dokument [4] beschrieben.

20

- Erkennung von Schlüsselwortkombinationen

Zur Erkennung von Schlüsselwortkombinationen wird das gleiche Prinzip wie zur Erkennung von Schlüsselworten verwendet, mit dem Unterschied, daß vordefinierte Wortketten in einem beliebigen Eingabesignal ES detektiert werden können. Die Wortketten  
25 können entweder explizit als Referenzdaten abgespeichert werden oder durch andere Verfahren, beispielsweise Language-Modelle oder Grammatiken beschrieben werden.

30 - Alphabeterkennung

Ist ein Spracherkennungsmodul Ei als eine Alphabeterkennungseinrichtung ausgestaltet, so werden üblicherweise die Buchstaben mit besonderen akustisch phonetischen Modellen beschrieben. Als lexikalische Einheiten, die in einem Speicher  
35 des Sprachverarbeitungssystems vorgesehen sind, werden in diesem Fall im wesentlichen die Buchstaben des Alphabets verwendet. Ein Alphabeterkenner kann beispielsweise als Einzel-

worterkenner das isolierte Alphabet oder auch als kontinuierlicher Erkenner für eine fließende Buchstabierung realisiert sein. Bei kontinuierlichen Alphabeterkennern kann durch Language-Modelle oder auch durch Grammatiken die Suche der tatsächlichen Wortfolgen unterstützt werden.

- Lautfolgeerkennung

Ein Spracherkennungsmodul Ei zur Lautfolgenerkennung zeichnet sich dadurch aus, daß die akustischen Einheiten auf phonetische Einheiten modelliert werden und die zu erkennenden Lautfolgen nicht durch ein Lexikon beschränkt sind. Lautfolgengerkenner bzw. Phonemerkenner beruhen auf akustisch-phonetischen Modellierungen der Laute einer Sprache. Mit dem phonetischen Inventar wird die Suche in dem akustischem Sprachsignal durchgeführt. Hypothetisiert wird eine beliebige Kette von Lauten. In der akustischen Analyse werden verschiedene Heuristiken verwendet, um ein zu schnelles Wechseln der hypothetisierten Laute, den Phonemen, in einer Äußerung zu verhindern. Zusätzlich können als Einschränkungen der zu erkennenden Lautfolgen im Sinne von Language-Modellen phonetische Gesetzmäßigkeiten einer Sprache berücksichtigt werden.

- DTMF-Erkennung

Eine DTMF-Erkennung ist aus verschiedensten Bereichen der telefonischen Spracherkennung bekannt.

Als Sprachausgabemodule Aj können verschiedene bekannte Prinzipien zur Realisierung verschiedener Sprachausgabeprinzipien verwendet werden:

30

- Ausgabe vorgegebener, gespeicherter Sprachkonserven

Ist ein Sprachausgabemodul Aj zur Ausgabe vorgegebener, gespeicherter Sprachkonserven ausgestaltet, so werden Sprachäußerungen, die vorher von einer beliebigen Person eingesprochen, aufgenommen und digitalisiert abgespeichert wurden, über einen beliebigen akustischen Kanal ausgegeben. Die aufgenommenen Äußerungen können hierbei die kompletten auszuge-

35

benden Äußerungen sein, die jeweils für eine bestimmte Situation vorgesehen sind, oder aber auch Bruckstücke davon, die zur Laufzeit wieder zu kompletten Äußerungen zusammengesetzt werden. Ein typisches Beispiel hierfür ist die Bildung von  
5 Ziffernketten, deren Länge und Zusammensetzung in den meisten Anwendungsfällen nicht vorhersehbar ist. Es können jedoch je nach Kontext verschiedene intonatorische Varianten ein und derselben Äußerung vorliegen und in einem Softwareprogramm entsprechend dem aktuellen Kontext ausgewählt werden. Ferner  
10 sind verschiedene Methoden anwendbar, die eine sog. akustische Glättung der verwendeten Konserven an den Übergängen der Konkatenation der einzelnen Sprachkonserven in verschiedenster Form vornehmen können.

15 - Ausgabe von aus gespeicherten Phonemen synthetisierten Worten

Mögliche Realisierungen zur Ausgestaltung von Sprachausgabemodulen zur Synthese von gespeicherten Phonemen zu synthetisierten Worten sind in dem Dokument [5] beschrieben.

20

Das Mittel MA zur Auswahl der in einem Anwendungsszenario jeweils zu verwendenden Spracherkennungsmodule Ei bzw. Sprachausgabemodule Aj ist derart ausgestaltet, daß abhängig von einem dem Mittel MA zugeführten Eingabesignal ES mindestens  
25 eines der Spracherkennungsmodule bzw. der Sprachausgabemodule aktiviert wird und, soweit es erforderlich ist, den einzelnen Spracherkennungsmodulen Ei bzw. Sprachausgabemodulen Aj Parameter, die im weiteren Verfahren zur Spracherkennung bzw. zur Sprachausgabe erforderlich sind, zugeführt werden. Die entsprechenden Parameter können beispielsweise darin zu sehen  
30 sein, daß den Spracherkennungsmodulen Ei bzw. den Sprachausgabemodulen Aj Ressourcen zugewiesen werden oder auch beispielsweise die Speicheradressen der verwendeten digitalen Lexika zugeführt werden.

35

In diesem Ausführungsbeispiel werden die Ergebnisse, also die ausgegebenen Sprachsignale aus den Sprachausgabemodulen Aj

dem Mittel MA zugeführt, und beispielsweise über einen Ausgang AS des Sprachverarbeitungssystems dem Benutzer B ausgegeben. Ein von einem Benutzer B gesprochenes Sprachsignal wird über einen Eingang E, beispielsweise realisiert durch  
5 einen Mikrophon, dem Mittel MA zugeführt, und über das Mittel MA abhängig von den jeweils aktivierten Spracherkennungsmodulen Ei, den jeweils aktivierten Spracherkennungsmodul Ei zugeführt.

- 10 Die einzelnen Module Ei, Aj verwenden u. a. Wissensquellen, beispielsweise digitale Lexika, allgemeine Daten, die in einem Speicher des Sprachverarbeitungssystems gespeichert sind. Unter den Wissensquellen sind beispielsweise auch auf die jeweilige Spracherkennungsaufgabe spezialisierte akustischer  
15 Modelle, z. B. ein Einzelworterkenner mit Lautmodellen oder Einzelworterkenner mit akustischen Modellen für beliebige Wörter, zu verstehen. Die Wissensquellen können aber auch die von den Erkennern für einen gegebenen Dialogschritt in einem Dialogsystem zur Auswahl stehenden spezifischen Wortschätze  
20 sein, oder für bestimmte, erwartete eingabespezifische sog. Sprachmodelle.

- Das von den Spracherkennungsmodulen Ei verwendete Vokabular ist beliebig durch orthographische, d. h. sprecherunabhängige, oder sprachliche, d. h. sprecherabhängige, Eingabe sowohl  
25 bei der Erstellung als auch aktuell zur Laufzeit des Sprachverarbeitungssystems erweiterbar und damit an wechselnde Anforderungen anpaßbar.

- 30 Durch geeignete Lernverfahren können die Wissensquellen während des Betriebs des Sprachverarbeitungssystems durch eine Auswertung der laufenden Eingabedaten modifiziert und der besonderen Nutzungs- und Anwendungssituation angepaßt werden.

- 35 Weiterhin kann ein Parameter, der den Spracherkennungsmodulen Ei oder den Sprachausgabemodulen Aj zugeführt wird darin zu sehen sein, daß mit dem Parameter angegeben wird, ob selbst



während einer Sprachausgabe von dem Benutzer B gesprochene Sprachsignale bearbeitet und von dem Sprachverarbeitungssystem aufgenommen werden können.

- 5 Im folgenden wird anhand eines sehr einfachen Beispiels für eine Dialogsituation eine Ausgestaltung des Sprachverarbeitungssystems als Sprachdialogsystem dargestellt.

10 In der folgenden Darstellung in einem Pseudo-Nachrichtenflußdiagramm wird jeweils zu Beginn einer Zeile auf der linken Seite eines Pfeils der Sender eines bestimmten Signals und auf der rechten Seite eines Pfeils der Empfänger des Signals, das im weiteren in dieser Zeile angegeben ist, beschrieben.

15

In diesem Beispiel wird ohne Einschränkung der Allgemeinheit von einer Weiterbildung, welche im weiteren beschrieben wird, ausgegangen, nämlich einer Dialogablaufsteuerung DA.

20 DA → MA: GebeAus("Bitte nennen Sie Ihre Nummer nach dem Piepton")

MA → A1: PlayPromptFile(Bitte\_nennen\_Sie\_Ihre\_Nummer\_nach\_dem\_Piepton)

A1 → MA: PlayPromptFinished

25 MA → DA: AusgabeComplete

DA → MA: ErkenneÄußerung(Domäne=Ziffern, Anzahl max.=5)

MA → E1: StarteErkennung(Modus=Einzelziffer, Lexikon=Digits)

30 MA → E2: StarteErkennung(Modus=kontinuierliche Ziffernerkennung, Lexikon= Digits)

E1 → MA: ErkenneAusgabe(Modus=Einzelziffer, Resultat=0)

E2 → MA: ErkenneAusgabe(Modus=kontinuierliche Ziffern, erkannte Ziffernfolge: 1 2 3 4 5)

35

MA → DA: EingabeZiffernfolge: (1 2 3 4 5)

DA → MA: GebeAus("Ihre Nummer ist 1 2 3 4 5")  
DA → A1: PlayPromptFile(Ihre\_Nummer\_ist)  
MA → A2: Synthesise(Ziffernfolge 1 2 3 4 5)  
A1 → MA: PlayPromptFinished  
5 A2 → MA: Synthesise Finished  
MA → DA: AusgabeComplete

DA → MA: GebeAus("Möchten Sie Vertrieb oder Kundenservice  
sprechen?")  
10 MA → A1: PlayPromptFile(Möchten\_Sie\_Vertrieb\_oder\_Kun-  
denservice-sprechen)  
A1 → MA: Prompt file not available  
MA → A2: Synthesise(Möchten Sie Vertrieb oder Kundenservice  
sprechen?)  
15 A2 → MA: Synthesis finished  
MA → DA: Ausgabe complete

DA → DM: ErkenneSprache(Domäne=Vertrieb\_oder\_Kundenservice)  
MA → E3: StarteErkennung(Modus=kontinuierlich, Lexi-  
20 kon=Kernvokabular)  
DM → E4: StarteErkennung(Modus=Schlüsselworterkennung, Lexi-  
kon=Vertrieb und Kundenservice)  
E3 → MA: ErkennerAusgabe(Modus=kontinuierlich, Resultat="den  
Vertrieb bitte")  
25 E4 → MA: ErkennerAusgabe(Modus=Schlüsselworterkennung, Re-  
sultat="Vertrieb")  
MA → DA: Eingabezeichenfolge(Vertrieb)

DA → MA: GebeAus("Ihr Geburtsdatum bitte")  
30 MA → A1: PlayPromptFile(Ihr\_Geburtsdatum\_bitte)

DA → MA: ErkenneÄußerung(Domäne=Datum)  
DM → E5: StarteErkennung(Modus=kontinuierlich, Lexi-  
kon=Datumsangabe)  
35 E5 → MA: ErkenneAusgabe(Modus=kontinuierlich, Resultat 1="am  
ersten ersten neunzehnhundertfünfzig", Resultat 2="am ersten  
elften neunzehnhundertfünfzig")

MA → DA: Eingabezeichenfolge(am ersten ersten neunzehnhundertfünfzig)  
DA → MA: GebeAus("Sie sind am ersten ersten neunzehnhundertfünfzig geboren")  
5 MA → A3: ConcatAusgabe(Sie\_sind\_am ersten, ersten, neunzehnhundertfünfzig, geboren)  
A3 → MA: PlayPromptConcatComplete  
MA → DA: AusgabeComplete  
DA → MA: ErkenneSprache(Domäne=j/n-Frage)  
10 MA → E6: StarteErkennung(Modus=Einzelwort, Lexikon=(ja, nein))

In diesem einfachen Beispiel weisen die Spracherkennungsmodulle Ei die folgenden einfachen Spracherkennungscharakteristika  
15 auf:

Das Spracherkennungsmodul E1 ist als Einzelziffernerkenner ausgestaltet, E2 als Spracherkennungsmodul zur Erkennung kontinuierlicher Ziffern, also von Ziffernketten. Das Spracherkennungsmodul E3 ist als Spracherkennungsmodul zur Erkennung  
20 fließend gesprochener Sprache mit einem begrenzten Vokabular, dem Lexikon „Kernvokabular“, ausgestaltet. Das Spracherkennungsmodul E4 ist als Schlüsselworterkenner mit dem Lexikon "Vertrieb und Kundenservice" ausgestaltet, das Spracherkennungsmodul E5 ist als ein Spracherkennungsmodul zur Erkennung  
25 von vorgegebenen Wortkombinationen, in diesem Spezialfall zur Erkennung von einer Datumsangabe ausgestaltet.

Die verwendeten Sprachausgabemodule Aj sind in diesem einfachen Beispiel als Modul zur Ausgabe vorgegebener, gespeicherter Sprachkonserven (A1), ein Modul zur Synthetisierung von  
30 Sprache aus beispielsweise Phonemen (A2) sowie ein Modul zur Ausgabe zusammengesetzter einzelner vorgegebener, gespeicherter Sprachkonserven (A3) ausgestaltet.

35 Wie aus diesem einfachen Beispiel ersichtlich wird, wird jeweils dem Mittel MA von den Spracherkennungsmodulen Ei die

Ergebnisse der Spracherkennung bzw. von den Sprachausgabemodulen Aj das auszugebende Sprachsignal zugeführt.

- 5 Andererseits wird von dem Mittel MA den einzelnen Modulen Ei, Aj die jeweilige zur Erkennung benötigte Information zugeführt, beispielsweise auch das in das Sprachverarbeitungssystem eingesprochene Sprachsignal.

- 10 Eine weitere Weiterbildung des Sprachverarbeitungssystem und des Verfahrens zur Sprachverarbeitung ist darin zu sehen, daß das Sprachsignal auch von mehreren Spracherkennungsmodulen Ei oder Sprachausgabemodulen Aj gleichzeitig bearbeitet wird, und daß jeweils das bezüglich eines beliebigen Ähnlichkeitsmaßes beste Ergebnis verwendet wird oder auch nur ein Vergleichsergebnis zur Erhöhung der Erkennungssicherheit verwendet wird.

#### Dialogablaufsteuerung

- 20 Mit der Dialogablaufsteuerung DA wird ein vorgebbare Dialog mit dem Benutzer B des Sprachverarbeitungssystems je nach der Anwendungssituation durchgeführt, und abhängig von der jeweiligen Dialogssituation dem Mittel MA Steuerungsinformation, beispielsweise noch benötigte Information zugeführt. Weiter-
- 25 hin wird in dieser Weiterbildung beispielsweise das Ergebnis der Spracherkennungsmodule in dem Mittel zur Dialogablaufsteuerung ausgewertet, was beispielsweise auf die folgende Weise erfolgen kann:

- 30 Die Dialogablaufsteuerung DA weist mindestens folgende Komponenten auf, die miteinander gekoppelt sind, beispielsweise über einen Bus BU (vgl. Figur 2):
- ein Eingabemittel EM,
  - einen Eingabespeicher ESP,
  - 35 - eine Aktionsermittlungseinheit AE,
  - einen Parameterparser PP.

Die Arbeitsweise der einzelnen Komponenten der Dialogablaufsteuerung DA werden im Zusammenhang mit den Verfahren zur Ablaufsteuerung eines Dialogs mit dem Benutzer B im weiteren detailliert erläutert.

5

Die Aktionsermittlungseinheit AE dient zum einen zur Ermittlung von Aktionshinweisen und zum anderen zur Ermittlung mindestens einer vorgebbaren Aktion, auf die durch den Aktionshinweis hingewiesen wird. Die Aktion wird aus einer beliebigen Anzahl vorgegebbarer Aktionen ermittelt. Verschiedene Verfahrenswesen zur Ermittlung der gesuchten Aktion, auf die durch den Aktionshinweis hingewiesen wird, sind dem Fachmann bekannt. Einige beispielhaft dargelegte Möglichkeiten zu deren Realisierung werden jedoch im Zusammenhang mit dem Verfahren detailliert erläutert.

10

15

Der Parameterparser PP dient zur Ermittlung von Aktionsparametern, die jeder Aktion jeweils eindeutig zugeordnet sind, aus der in dem Eingabespeicher ESP gespeicherten digitalen Aktionsinformation.

20

In Figur 2 ist die Dialogablaufsteuerung DA mit einigen Weiterbildungen der Dialogablaufsteuerung DA zur Aktionsermittlung dargestellt.

25

Eine in einer Ausgestaltung vorgesehene Informationsdatenbank ID enthält jeweils anwendungsspezifisch gesuchte Information, die anhand der bestimmten Aktion und den anschließend für die Aktion ermittelten Aktionsparameter ermittelt wird.

30

Es ist in einer Weiterbildung eine Aktionsklärungseinheit AKE vorgesehen zur Ermittlung weiterer Aktionshinweise von dem Benutzer. Die Aktionsklärungseinheit AKE gewinnt vor allem in Anwendungsgebieten eine erhöhte Bedeutung, wenn es gilt, unter verschiedenen Aktionen die gesuchte Aktion zu ermitteln. Sind die Aktionshinweise, die von dem Benutzer B der Dialogablaufsteuerung DA übergeben werden, zur Bestimmung der Ak-

35

tion nicht ausreichend, so wird die Aktionsklärungseinheit AKE zur Ermittlung weiterer Aktionshinweise von dem Benutzer B verwendet. Die Ausgestaltung der Aktionsklärungseinheit AKE wird im Zusammenhang mit dem Verfahren näher erläutert. Es ist jedoch nicht notwendig, daß die Aktionsklärungseinheit AKE genau eine Aktion ermittelt. Es ist ebenso vorgesehen, die Anzahl vorgegebener Aktionen in einer beliebigen Art zu reduzieren, um somit nur noch eine geringere Anzahl von Aktionen weiter zu verarbeiten.

10

In einer Weiterbildung ist eine Parameterklärungseinheit PKE vorgesehen zur Ermittlung weiterer Aktionsparameter von dem Benutzer B. Die Parameterklärungseinheit PKE wird verwendet, wenn für die mindestens eine ermittelte Aktion nicht alle Aktionsparameter, die den jeweiligen Aktionen zugeordnet sind, bestimmt werden können. Für diesen Fall ist es vorteilhaft, daß die Parameterklärungseinheit PKE verwendet wird, um fehlende Aktionsparameter von dem Benutzer B zu ermitteln.

15

20 In einer Weiterbildung der Dialogablaufsteuerung DA ist es ferner vorteilhaft, einen ersten Aktionsspeicher AS1 und/oder einen zweiten Aktionsspeicher AS2 vorzusehen. In dem ersten Aktionsspeicher AS1 werden Aktionen und deren Aktionsparameter gespeichert, bei denen der mindestens eine Aktionshinweis vollständig mit mindestens einem Teil von Schlüsselbegriffen, deren Bedeutung im weiteren beschrieben wird, übereinstimmt. In dem zweiten Aktionsspeicher AS2 werden Aktionen und deren Aktionsparameter gespeichert, bei denen mindestens einer der Aktionshinweise mit mindestens einem Schlüsselbegriff übereinstimmt.

25

30

Ferner ist in einer Weiterbildung ein Mittel GF zur Generierung von Fragen (prompts) an den Benutzer B zur Ermittlung weiterer Aktionshinweise und/oder weiterer Aktionsparameter vorgesehen. Der Aufbau des Mittels GF zur Generierung von Fragen ist beispielsweise aus dem Dokument [3] bekannt.

35

Ferner ist es in einer Weiterbildung der Dialogablaufsteuerung DA vorgesehen, eine Steuerungseinheit STE zur Steuerung einer Hintergrundanwendung HA unter Verwendung der ermittelten Information zu verwenden. Der Aufbau der Steuerungseinheit STE ist völlig anwendungsspezifisch und dient lediglich dazu, die ermittelte Aktion zur Steuerung der Hintergrundanwendung HA zu verwenden. Die unterschiedlichen Verwendungsmöglichkeiten und somit auch die unterschiedlichen Ausgestaltungen der Steuerungseinheit STE werden im weiteren detailliert beschrieben.

Ferner ist in einer Weiterbildung ein Schlüsselbegriffeditor SE vorgesehen, mit dem zum einen die im weiteren beschriebenen Schlüsselbegriffe und zum anderen neue Aktionen und/oder Aktionsparameter verändert, entfernt oder hinzugefügt werden können.

Die einzelnen Verfahrensschritte des Verfahrens sind in Figur 3 dargestellt.

In einem ersten Schritt 301 wird von dem Benutzer B der Dialogablaufsteuerung DA Aktionsinformation, die mindestens einen Aktionshinweis aufweist, übergeben. Das Verfahren muß nicht notwendigerweise mit der im vorigen beschriebenen Anwendung durchgeführt werden, es ist ebenso möglich, das Verfahren mit Hilfe eines Rechners durchzuführen.

Die Aktionsinformation wird von dem Benutzer B an die Dialogablaufsteuerung DA oder das Sprachverarbeitungssystem übergeben. Die Aktionsinformation wird digitalisiert und als digitalisierte Aktionsinformation gespeichert 302. Die digitale Aktionsinformation kann beispielsweise in einer Phonemdarstellung oder auch in Form eines üblichen Codes zur Darstellung von Schriftzeichen (beispielsweise ASCII-Code) gespeichert werden.

In einem weiteren Schritt 303 wird aus der gespeicherten Aktionsinformation unter Verwendung des Parameterparses PP, dessen Aufbau beispielsweise aus dem Dokument [4] bekannt ist, mindestens ein Aktionshinweises ermittelt.

5

Unter Verwendung des Aktionshinweises und der Aktionsermittlungseinheit AE wird mindestens eine Aktion bestimmt 304.

10

Die Bestimmung der mindestens einen Aktion kann auf unterschiedliche Weise erfolgen. Eine sehr einfach jedoch in keiner Weise ausschließlich verwendbare Methode liegt darin, jeder Aktion eine vorgebbare Anzahl von Schlüsselbegriffen, die jeweils die Aktion charakterisieren, zuzuordnen, und die aus der Aktionsinformation ermittelten Aktionshinweise, die

15

durch den Parameterparser PP ermittelt werden, mit den Schlüsselbegriffen zu vergleichen. Der Vergleich kann durch einen direkten Wortvergleich oder auch durch einen beliebigen Mustervergleich, deren Algorithmen jedem Fachmann geläufig sind, erfolgen.

20

Ist die mindestens eine Aktion ermittelt worden 304, werden für alle Aktionen die Aktionsparameter, die den Aktionen eindeutig zugeordnet sind, aus der digitalen Aktionsinformation mittels des Parameterparser PP ermittelt 305. Anhand der ermittelten Aktion sowie der zugehörigen Aktionsparameter wird

25

nun in einem letzten Schritt 306 die Aktion durchgeführt.

30

In einer Weiterbildung des Verfahrens wird Information zu der Aktion aus der Informationsdatenbank ID ermittelt und dem Benutzer B über die Ausgabeeinheit AUS dargestellt.

35

Reichen die in der digitalen Aktionsinformation enthaltenen Aktionshinweise nicht für die Ermittlung mindestens einer Aktion aus 401, so ist es in einer Weiterbildung vorgesehen, weitere Aktionshinweise beispielsweise durch einen Dialog mit dem Benutzer B zu ermitteln 402.



Dies geschieht z. B. unter Verwendung der Aktionsklärungseinheiten AKE, die vorzugsweise derart ausgestaltet ist, daß für die vorgegebenen Aktionen jeweils mindestens ein Ermittlungsprompt eindeutig zugeordnet ist. Unter dem Ermittlungsprompt sind beispielsweise Fragen zu verstehen, die der jeweiligen Aktion zugeordnet und fest gespeichert sind. Diese Ermittlungsprompts werden dem Benutzer B bei unzureichender Information zur Ermittlung der Aktionen dargestellt. Der Benutzer B gibt auf diese Ermittlungsprompts, also die entsprechenden Fragen, Antworten, die dann wiederum verarbeitet werden, wodurch weitere Aktionshinweise ermittelt werden. Der Dialog wird solange durchgeführt und wiederholt, bis die ermittelten Aktionshinweise und weiteren Aktionshinweise ausreichend sind zur Identifikation der vorgebbaren Anzahl von Aktionen.

Ist die mindestens eine Aktion ermittelt, werden für jede Aktion die Aktionsparameter, die der jeweiligen Aktion zugeordnet sind, ermittelt 305.

Sind jedoch nicht alle benötigten Aktionsparameter vollständig aus der in dem Eingabespeicher ESP ermittelten digitalen Aktionsinformation bekannt bzw. ermittelt worden, was in einem zusätzlichen Verfahrensschritt 403 geprüft wird, werden die restlichen Aktionsparameter beispielsweise von der Parameterklärungseinheit PKE ermittelt 404.

Dies geschieht beispielsweise ebenfalls in einem Dialog mit dem Benutzer B. Für diesen Dialog können wiederum den einzelnen Parametern zugeordnete Prompts, im folgenden als Parameter-Prompts bezeichnet, verwendet werden, die bei notwendiger Ermittlung der entsprechenden Aktionsparameter an den Benutzer B ausgegeben werden und der Benutzer B antwortet auf die entsprechenden Fragen.

Sowohl die Ermittlung weiterer Aktionshinweise 402 als auch die Ermittlung der restlichen Aktionsparameter 404 kann jedoch auch auf andere Weise erfolgen.

Beispielsweise ist es nicht unbedingt notwendig, fest vorgegebene Ermittlungs-Prompts oder Parameter-Prompts den einzelnen Aktionen bzw. Parametern zuzuordnen, wodurch zum einen zwar die Einfachheit in der Durchführung des Verfahrens und in der Anpassung der jeweiligen Anwendung und die Erweiterbarkeit der Aktionen und Parameter erheblich vereinfacht wird, jedoch immer nur fest vorgegebene Fragen an den Benutzer B gestellt werden.

Es ist in einer Weiterbildung des Verfahrens und der Dialogablaufsteuerung DA vorgesehen, das Mittel GF zur Generierung von Fragen an den Benutzer B zu verwenden. Eine Möglichkeit zur Realisierung des Mittels GF ist beispielsweise in dem Dokument [3] beschrieben.

In einer Weiterbildung des Verfahrens ist vorgesehen, zur Ermittlung der Aktion in der Aktionsermittlungseinheit AE folgende Schritte durchzuführen. In der Aktionsermittlungseinheit AE wird für jede Aktion der Vergleich der Schlüsselbegriffe mit den Aktionshinweisen durchgeführt.

Es wird eine erste Menge von Aktionen ermittelt, bei denen alle Aktionshinweise mit mindestens einem Teil der Schlüsselbegriffe übereinstimmen. Anschaulich beschrieben bedeutet diese Vorgehensweise, daß die Schnittmenge aller Aktionen, auf die durch die Aktionshinweise hingewiesen wurde, ermittelt wird. Die erste Menge von Aktionen wird beispielsweise in dem ersten Aktionsspeicher AS1 gespeichert.

Ferner wird eine zweite Menge von Aktionen gebildet, bei denen mindestens ein Teil der Aktionshinweise mit den Schlüsselbegriffen übereinstimmt. Diese Vorgehensweise bedeutet anschaulich eine Bildung der Vereinigungsmenge aller Aktionen, auf die hingewiesen wurde. Die zweite Menge von Aktionen wird beispielsweise in dem zweiten Aktionsspeicher AS2 gespeichert.

Enthält die erste Menge genau ein Element, so ist die zu ermittelnde Aktion eindeutig identifiziert. Für diesen Fall ist kein weiterer Klärungsdialog notwendig. Enthält die erste Menge jedoch mehr Elemente als die Anzahl von zu ermittelnden Aktionen, so ist die Aktion durch die initiale Benutzereingabe nicht genau genug beschrieben. In diesem Fall ist es vorteilhaft, den im vorigen beschriebenen Klärungsdialog durchzuführen.

10 Ist die erste Menge eine leere Menge, so enthält die initiale Benutzeräußerung Hinweise auf disjunkte Aktionen. In diesem Fall wird vorteilhafterweise der Klärungsdialog durch die in der zweiten Menge enthaltenen Aktionen gesteuert.

15 Sind sowohl die erste Menge als auch die zweite Menge leere Mengen, so wird beispielsweise ein Klärungsdialog zur Unterscheidung aller in der Anwendung bekannten Aktionen durchgeführt.

20 Somit dienen die erste Menge und die zweite Menge in dieser Weiterbildung als Grundlage für den Dialog mit dem Benutzer B zur Ermittlung der mindestens einen Aktion.

25 Die ermittelte Aktion kann in einer Weiterbildung des Verfahrens vorteilhaft zur Steuerung mindestens einer Hintergrundanwendung HA eingesetzt werden.

30 Unter der Hintergrundanwendung HA sind unterschiedlichste Einsatzmöglichkeiten des Verfahrens und der Dialogablaufsteuerung DA zu verstehen, die im weiteren erläutert werden.

Das Verfahren sowie die Dialogablaufsteuerung DA können beispielsweise in folgenden Gebieten vorteilhaft eingesetzt werden:

- in einem Telefondienst,
- in einem Telekauf-System und/oder Tele-Banking-System,

- in einem sog. Voice Mail Handling System,
- zur Steuerung einer beliebigen Maschine, beispielsweise einer Werkzeugmaschine oder eines Roboters,
- zur Steuerung eines Rechners,
- 5 - in einem Messaging-System einer Nebenstellenanlage eines Telekommunikationssystems.

Ferner ist es in einer Weiterbildung des Verfahrens vorteilhaft, daß die Auswertung der digitalen Aktionsinformation, die in dem Eingabespeicher ESP gespeichert wird, von dem Parameterparsel PP in bezüglich der Eingabe der Aktionsinformation in umgekehrter Reihenfolge ermittelt werden.

Dies bedeutet, daß die Auswertung der Aktionsinformation am Schluß der Aktionsinformation begonnen wird und bis zum Anfang der Aktionsinformation weitergeführt wird. Durch diese Vorgehensweise wird eine Korrektur der Aktionsinformation durch den Benutzer B in einem Satz möglich, was im folgenden noch detailliert erläutert wird.

Zum einfacheren Verständnis der Dialogablaufsteuerung DA wird im folgenden ein sehr einfaches Beispiel aus der Touristikbranche dargestellt. Beispielsweise können in diesem Anwendungsgebiet folgende Aktionen definiert sein:

- Auskunft über Bahnverbindungen;
- Auskunft über Flugverbindungen;
- Auskunft über Schiffsverbindungen;
- Auskunft über Hotelzimmer;
- 30 - Restaurantinformation.

Die einzelnen Aktionen können beispielsweise durch folgende Parameter gekennzeichnet sein;

- Bahnauskunft: <Abfahrtsort>, <Ankunftsart>, <Datum>, <Uhrzeit>
- 35 - Flugauskunft: <Abfahrtsort>, <Ankunftsart>, <Datum>, <Uhrzeit>

- Schiffsauskunft: <Abfahrtsort>, <Ankunftsart>, <Datum>, <Uhrzeit>
- Hotelauskunft: <Ort>, <Zimmer>, <Art>, <Anzahl der Personen>, <Check in Datum>, <Check out Datum>
- Restaurantinformation: <Ort>, <Stil>, <Kategorie>.

In Form eines Pseudocodes werden im folgenden die einzelnen Aktionen spezifiziert. Hierbei wird jeweils unter dem Punkt 2.1 der Aktion eindeutig eine Nummer zugewiesen, unter dem Punkt 2.2 der jeweilige Ermittlungs-Prompt und unter der Nummer 2.3 die einzelnen Parameterspezifikationen, die sich, jeweils abhängig von der definierten Grammatik in diesem Fall für vorgegebene Parameter-Prompts in einer Weise darstellen, daß jeweils unter dem Punkt 3.1 und dem Punkt 3.3 die jeweilige Grammatik des jeweiligen Aktionsparameters gegeben ist und unter dem Punkt 3.2 der jeweilige Parameter-Prompt dargestellt ist.

20

aktion (Bahnauskunft)

{2.1: Nummer = 1

2.2: Ermittlungs-prompt = 'Möchten Sie eine Bahnauskunft einholen?'

2.3: Parameterspezifikationen = {parameter

{3.1: g\_abfahrt(X)

3.2: 'Bitte nennen Sie

den gewünschten Abfahrtsort.'

3.3: g\_ort(X)

}

30

parameter

{3.1: g\_ankunft(Y)

3.2: 'Bitte nennen Sie

den gewünschten Ankunftsart.'

3.3: g\_ort(Y)

35

}

parameter

```

                                {3.1: g_datum(Z)
                                3.2: 'An welchem Datum
wollen Sie fahren?'
                                3.3: g_datum(Z)
                                }
5      parameter
      {3.1: g_uhrzeit(W)
      3.2: 'Um wieviel Uhr wol-
len Sie fahren?'
      3.3: g_uhrzeit(W)
10     }
    }
    2.4: Aktionshinweise = {von, nach, fahren, Zugauskunft,
Eisenbahn, Bahnauskunft}
15    2.5: Bestätigungs-prompt =
        'Sie erhalten Auskunft über die Züge von X
nach Y am Z um W Uhr.'
    }

20  aktion (Flugauskunft)
    {2.1: Nummer = 2
    2.2: Ermittlungs-prompt = 'Möchten Sie eine Flugauskunft
einholen?'
25    2.3: Parameterspezifikationen = {parameter
                                {3.1: g_abfahrt(X)
                                3.2: 'Bitte nennen Sie
den gewünschten Abflugort.'
                                3.3: g_ort(X)
                                }
30      parameter
      {3.1: g_ankunft(Y)
      3.2: 'Bitte nennen Sie
den gewünschten Ankunftsart.'
      3.3: g_ort(Y)
35     }
    }
    parameter
```

```

{3.1: g_datum(Z)
  3.2: 'An welchem Datum
wollen Sie fliegen?'
  3.3: g_datum(Z)
}
5 parameter
  {3.1: g_uhrzeit(W)
    3.2: 'Um wieviel Uhr wol-
len Sie fliegen?'
    3.3: g_uhrzeit(W)
  }
10
}
2.4: Aktionshinweise = {von, nach, fahren, fliegen,
Flugauskunft, Flugzeug}
15 2.5: Bestätigungs-prompt =
      'Sie erhalten Auskunft über die Flüge von X
nach Y am Z um W Uhr.'
}

20 aktion (Schiffsauskunft)
  {2.1: Nummer = 3
    2.2: Ermittlungs-prompt = 'Möchten Sie eine Schiffsaus-
kunft einholen?'
    2.3: Parameterspezifikationen = {parameter
25      {3.1: g_abfahrt(X)
        3.2: 'Bitte nennen Sie
den gewünschten Abfahrtsort.'
        3.3: g_ort(X)
      }
      parameter
        {3.1: g_ankunft(Y)
          3.2: 'Bitte nennen Sie
den gewünschten Ankunftsart.'
          3.3: g_ort(Y)
        }
      }
30
parameter
  {3.1: g_ankunft(Y)
    3.2: 'Bitte nennen Sie
den gewünschten Ankunftsart.'
    3.3: g_ort(Y)
  }
35
parameter
```

```

{3.1: g_datum(Z)
  3.2: 'An welchem Datum
wollen Sie fahren?'
  3.3: g_datum(Z)
}
5 parameter
{3.1: g_uhrzeit(W)
  3.2: 'Um wieviel Uhr wol-
len Sie fahren?'
  3.3: g_uhrzeit(W)
}
10
}
2.4: Aktionshinweise = {von, nach, fahren, Schiffsaus-
kunft, Schiff}
15 2.5: Bestätigungs-prompt =
'Sie erhalten Auskunft über die Schiffsverbindungen
von X nach Y am Z um W Uhr.'
}
20
aktion (Hotelauskunft)
{2.1: Nummer = 4
  2.2: Ermittlungs-prompt = 'Möchten Sie eine Hotelauskunft
einholen?'
25 2.3: Parameterspezifikationen = {parameter
{3.1: g_ortsangabe(X)
  3.2: 'In welchem Ort su-
chen Sie ein Hotel?'
  3.3: g_ortsangabe(X)
}
30 parameter
{3.1: g_zimmerart(Y)
  3.2: 'Suchen Sie Einzel-
zimmer oder Doppelzimmer?'
  3.3: g_zimmerart(Y)
35 }
parameter
```



31

```

{3.1: g_anzahl_personen(Z)
  3.2: 'Für wieviele Perso-
nen suchen Sie Zimmer?'
  3.3: g_anzahl_personen(Z)
}
5 parameter
{3.1: g_von_datum(W)
  3.2: 'Wann wollen Sie
einchecken?'
  3.3: g_datum(W)
}
10 parameter
{3.1: g_bis_datum(V)
  3.2: 'Wann wollen Sie au-
schecken?'
  3.3: g_datum(V)
}
15 }
2.4: Aktionshinweise = {Hotel, Unterkunft, Einzelzimmer,
20 Doppelzimmer, Zimmer}
2.5: Bestätigungs-prompt =
'Sie erhalten Auskunft über Hotels in X, die Y für Z
Personen vom W bis zum V frei haben.'
}
25
aktion (Restaurantinformation)
{2.1: Nummer = 5
  2.2: Ermittlungs-prompt = 'Möchten Sie eine Restaurantin-
30 formation einholen?'
  2.3: Parameterspezifikationen = {parameter
    {3.1: g_ortsangabe(X)
      3.2: 'In welchem Ort
möchten Sie essen gehen?'
      3.3: g_ort(X)
    }
  }
35 parameter
```

```

                    {3.1: g_stil(Y)
                    3.2: 'Bitte nennen Sie
den gewünschten Stil.'
                    3.3: g_stil(Y)
                    }
5      parameter
      {3.1: g_kategorie(Z)
      3.2: 'Bitte nennen Sie
die gewünschte Kategorie.'
      3.3: g_kategorie(Z)
10     }
    }
    2.4: Aktionshinweise = {Restaurant, essen}
    2.5: Bestätigungs-prompt =
15     'Sie erhalten Auskunft über Y Restaurants der Z Ka-
    tegorie in X.'
    }

```

20 In der folgenden Tabelle sind Schlüsselbegriffe für die einzelnen Aktionen, die durch die entsprechende Nummer in den geschweiften Klammern eindeutig identifiziert sind, beispielhaft aufgeführt.

	von	{1, 2, 3}
25	nach	{1, 2, 3}
	fahren	{1, 2, 3}
	Zugauskunft	{1}
	Eisenbahn	{1}
	fliegen	{2}
30	Flugauskunft	{2}
	Flugzeug	{2}
	Schiffsauskunft	{3}
	Schiff	{3}
	Hotel	{4}
35	Unterkunft	{5}
	Einzelzimmer	{4}
	Doppelzimmer	{4}

Restaurant {5}  
Essen {5}

5 Ein Beispiel für einen Dialog des Benutzers B mit der Dialogablaufsteuerung DA wird im folgenden dargestellt.

Der Benutzer B äußert seinen Wunsch präzise und vollständig.

Der Benutzer B gibt folgende Informationen ein:

10 „Ich möchte am 1. Januar 1996 um 5.00 Uhr mit dem Zug von München nach Hamburg fahren.“

15 Der erste Aktionsspeicher AS1 enthält unter Verwendung der oben beschriebenen Schlüsselbegriffe und Grammatiken die erste Menge von Aktionen {1}. Der zweite Aktionsspeicher AS2 enthält die zweite Menge von Aktionen {1, 2, 3}.

20 Da der erste Aktionsspeicher AS1 genau ein Element aufweist, wird direkt von der Dialogablaufsteuerung DA aus der Informationsdatenbank ID die gesuchte Information ermittelt und beispielsweise dem Benutzer B durch folgende Antwort dargestellt:

25 „Sie erhalten Auskunft über die Flüge von München nach Hamburg am 1. Januar 1996 um fünf Uhr: „1. Flug, 2. Flug, 3. Flug usw.“ .

Bei einem weiteren Beispiel nennt der Benutzer B weder Uhrzeit noch Verkehrsmittel:

30 Die Aktionsinformation ist beispielsweise folgende:  
„Ich möchte am 1. Januar 1996 von München nach Hamburg.“

35 Der erste Aktionsspeicher AS1 enthält in diesem Fall die erste Menge von Aktionen {1,2,3}. Der zweite Aktionsspeicher AS2 enthält die zweite Menge von Aktionen {1,2,3}. In der Dialogablaufsteuerung DA wird ein Klärungsdialog initiiert, um zwischen den Aktionen {1,2,3} zu unterscheiden.

Hierzu wird beispielsweise der oben beschriebene Ermittlungs-Prompt:

- 5 „Möchten Sie eine Bahnauskunft?“  
dem Benutzer B ausgegeben.

Der Benutzer B antwortet beispielsweise mit „Nein“.

- 10 In diesem Fall wird von der Dialogablaufsteuerung DA beispielsweise folgender Ermittlungsprompt ausgegeben:  
„Möchten Sie eine Flugauskunft?“

- 15 Beantwortet der Benutzer B diese Frage mit „Ja“, so wird ein weiterer Klärungsdialog zur Ermittlung der restlichen, noch fehlenden Aktionsparameter für die Flugauskunft durchgeführt. Hierbei wird beispielsweise zur Ermittlung des Aktionsparameters Uhrzeit der Parameter-Prompt:  
„Um wieviel Uhr möchten Sie fliegen?“  
dem Benutzer B ausgegeben.

- 20 Der Benutzer B antwortet beispielsweise:  
„Um fünf Uhr.“

- 25 Nun hat die Dialogablaufsteuerung DA ausreichend Information, um die gesuchte Aktion zu ermitteln. Das Ergebnis wird dem Benutzer B beispielsweise in folgender Weise dargestellt:  
„Sie erhalten Auskunft über die Flüge von München nach Hamburg am 1. Januar 1996 um fünf Uhr“.

- 30 Weitere Details über eine mögliche Ausgestaltung der Dialogablaufsteuerung DA sind in der Deutschen Patentanmeldung mit dem amtlichen Aktenzeichen 19615693.9 beschrieben.

- 35 In Fig. 4 ist in einem Ablaufdiagramm das Verfahren in seinen Verfahrensschritten dargestellt. Das Eingabesignal ES wird analysiert, wobei mit dem Eingabesignal ES beschrieben wird,

welche Art von Spracherkennung und/oder welche Art von Sprachausgabe im weiteren Verfahren benötigt wird 401.

5 Ferner wird abhängig von dem Eingabesignal ES mindestens ein Spracherkennungsmodul Ei und/oder ein Sprachausgabemodul Aj aktiviert 402.

10 In einem letzten Schritt 403 wird in mindestens einen der Spracherkennungsmodule und/oder der Sprachausgabemodule die jeweilige Spracherkennung bzw. Sprachausgabe durchgeführt.

15 Ferner ist es in einer Weiterbildung vorgesehen, gemeinsame Vorverarbeitungsschritte oder auch gemeinsame Teile von Algorithmen, die gemeinsam in verschiedenen Verfahren zur Spracherkennung gleichzeitig verwendet werden, in einem separaten Vorverarbeitungsmodul VV durchzuführen.

20 In der gemeinsamen Vorverarbeitung z. B. in mindestens einem Vorverarbeitungsmodul VV können beispielsweise eine akustische Vorverarbeitung oder eine sog. Abstandsberechnung der untersuchten Teile des Sprachsignals realisiert werden.

25 Ferner ist es in einer Weiterbildung vorteilhaft, Ressourcen gemeinsam zu nutzen, beispielsweise in einer Weise, daß ein Lexikon von mehreren Spracherkennungsmodulen Ei gemeinsam verwendet wird. Dies ist beispielsweise möglich, indem ein Lexikon zur Erkennung fließend gesprochener Sprache u. a. die Worte "Ja" und "Nein" enthält, wodurch es möglich ist, daß auch ein Schlüsselwort-Erkenner (Keyword-Spotter) auf die Begriffe "Ja" und "Nein" zur Durchführung der Spracherkennung  
30 in diesem Fall auf dasselbe Lexikon zugreifen kann.

Auch eine gemeinsame Nutzung von Ressourcen bei der Nachbearbeitung sind sowohl bei dem Verfahren als auch mit einem  
35 Nachbearbeitungsmodul bei dem Sprachverarbeitungssystem als eine Ausgestaltung vorgesehen.

5 Sowohl das Sprachverarbeitungssystem als auch das Verfahren zur Sprachverarbeitung kann sehr vorteilhaft sowohl in der Spracherkennung und in der Sprachausgabe sowie in der Durchführung eines Dialogs mit einem Benutzer B als auch zur Erstellung von Sprachdialogsystemen verwendet werden, da eine sehr einfache Zusammenstellung der anwendungsspezifischen Kombinationen von Spracherkennungsmodulen und/oder Sprachausgabemodulen möglich ist.

In diesem Dokument wurden folgende Veröffentlichungen zitiert:

- 5 [1] G. Ruske, Automatische Spracherkennung, Oldenbourg-Verlag, 1. Auflage, ISBN 3-486-20877-2, München, S. 106 - 171, 1988
- [2] A. Zell, Simulation neuronaler Netze, Edison Wesley, 1. Auflage, ISBN 3-89319-554-8, S. 137 - 143, 1994
- 10 [3] Caspari, R., 1993, Bottom Up Generation from Flat Semantic Structures, Proc. Deklarative und prozedurale Aspekte der Sprachverarbeitung, Deutsche Gesellschaft für Sprachverarbeitung, Sektion Computerlinguistik, 4. Fachtagung, S. 1 - 6
- 15 [4] Naumann, Sven und Hagen Langer, 1994: Parsing, Stuttgart, Teubner, 1. Auflage, ISBN 3-519-02139-0, S. 155 - 166
- 20 [5] DE 44 12 309 A1
- [6] GB 2 119 208 A

## Patentansprüche

1. Sprachverarbeitungssystem,
  - mit mehreren Spracherkennungsmodulen ( $E_i$ ,  $i=1..n$ ) und/oder Sprachausgabemodulen ( $A_j$ ,  $j=1..m$ ), die jeweils für eine vor-  
5 gebbare Art einer Spracherkennung und/oder einer Sprachausgabe vorgesehen sind, und
  - mit einem Mittel (MA) zur Auswahl mindestens eines Spracherkennungsmoduls ( $E_i$ ) und/oder Sprachausgabemoduls ( $A_j$ ) für  
10 eine im weiteren durchzuführende Spracherkennung und/oder Sprachausgabe, abhängig von einem Eingangssignal (ES), mit dem beschrieben wird, welche Art von Spracherkennung und/oder Sprachausgabe im weiteren benötigt wird.
- 15 2. Sprachverarbeitungssystem nach Anspruch 1,  
bei dem das Mittel (MA) zur Auswahl des Spracherkennungsmoduls ( $E_i$ ) und/oder des Sprachausgabemoduls ( $A_j$ ) derart ausgestaltet ist, daß das Spracherkennungsmodul ( $E_i$ ) und/oder das Sprachausgabemodul ( $A_j$ ) von dem Mittel (MA) gesteuert wird.
- 20 3. Sprachverarbeitungssystem nach Anspruch 1 oder 2,  
bei dem Teile von dem Spracherkennungsmodul ( $E_i$ ) und/oder dem Sprachausgabemodul ( $A_j$ ), die gemeinsam verwendet werden, in  
mindestens einem Vorverarbeitungsmodul (VV) und/oder in min-  
25 destens einem Nachbearbeitungsmodul realisiert sind.
4. Sprachverarbeitungssystem nach Anspruch 3,  
bei dem mehrere Spracherkennungsmodule ( $E_i$ ) und/oder Sprachausgabemodule ( $A_j$ ) gemeinsame Ressourcen verwenden.
- 30 5. Sprachverarbeitungssystem nach einem der Ansprüche 1 bis 4,  
bei dem eine Dialogablaufsteuerung (DA) vorgesehen ist, mit  
der ein Dialog des Sprachverarbeitungssystems mit einem Be-  
35 nutzer (B) des Sprachverarbeitungssystems realisiert wird.



6. Sprachverarbeitungssystem nach einem der Ansprüche 1 bis 5, bei dem das Eingangssignal (ES) von der Dialogablaufsteuerung (DA) erzeugt wird.

5 7. Sprachverarbeitungssystem nach einem der Ansprüche 1 bis 6,  
- bei dem das Spracherkennungsmodul (Ei) mindestens zur Durchführung einer der folgenden Arten der Spracherkennung  
10 ausgebildet ist:  
-- Einzelziffererkennung,  
-- Erkennung von Ziffernketten,  
-- Erkennung von Wörtern aus einem begrenzten Vokabular,  
-- Erkennung eines Einzelwortes mit einem unbegrenzten Voka-

15 bular,  
-- Erkennung fließend gesprochener Sprache mit einem unbegrenzten Vokabular,  
-- Erkennung von vorgegebenen Wortkombinationen,  
-- Schlüsselwort-Erkennung,  
20 -- Alphabeterkennung,  
-- Lautfolgernerkennung,  
-- Sprechererkennung,  
-- DTMF-Erkennung,  
und/oder

25 - bei dem das Sprachausgabemodul (Aj) mindestens zur Durchführung einer der folgenden Arten der Sprachausgabe ausgebildet ist:  
-- Ausgabe vorgegebener, gespeicherter Sprachkonserven,  
-- Ausgabe zusammengesetzter einzelner vorgegebener, gespeicherter Sprachkonserven,  
30 -- Ausgabe von aus gespeicherten Phonemen synthetisierten Worten,  
-- Ausgabe von DTMF-Tönen.

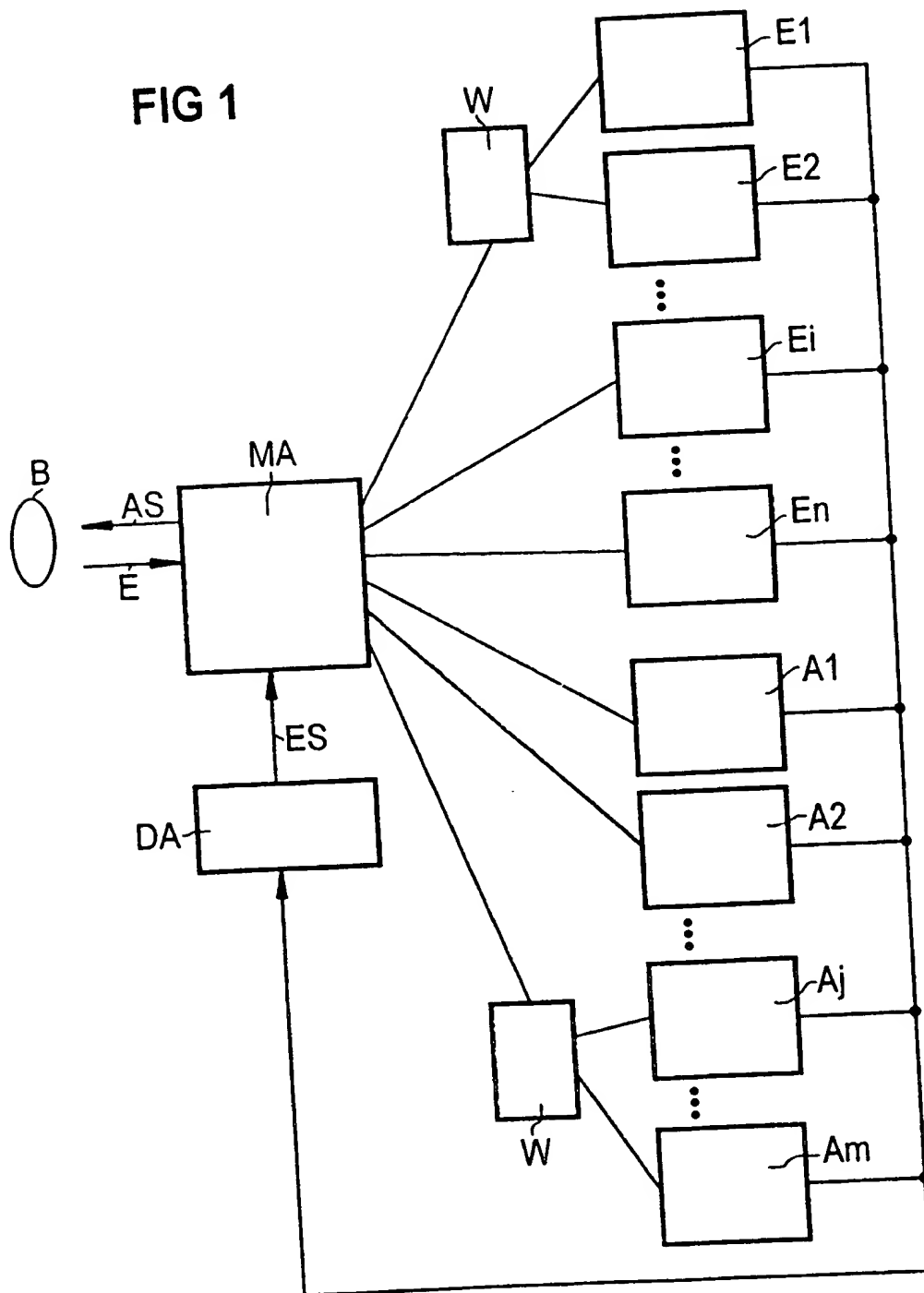
35 8. Verfahren zur Sprachverarbeitung,

- bei dem ein Eingangssignal (ES) analysiert wird (401), mit dem beschrieben wird, welche Art von Spracherkennung und/oder Sprachausgabe im weiteren benötigt wird,
- bei dem abhängig von dem Eingangssignal (ES) mindestens ein
- 5 Spracherkennungsmodul (Ei) und/oder mindestens ein Sprachausgabemodul (Aj), die jeweils für eine vorgebbare Art einer Spracherkennung und/oder einer Sprachausgabe vorgesehen sind, aktiviert wird (402), und
- bei dem mit dem ausgewählten Spracherkennungsmodul und/oder
- 10 dem Sprachausgabemodul die jeweilige Art der Spracherkennung und/oder der Sprachausgabe durchgeführt wird (403).

9. Verfahren nach Anspruch 8,  
bei dem vor der Spracherkennung eine Vorverarbeitung (VV) des  
15 zu erkennenden Sprachsignals erfolgt.

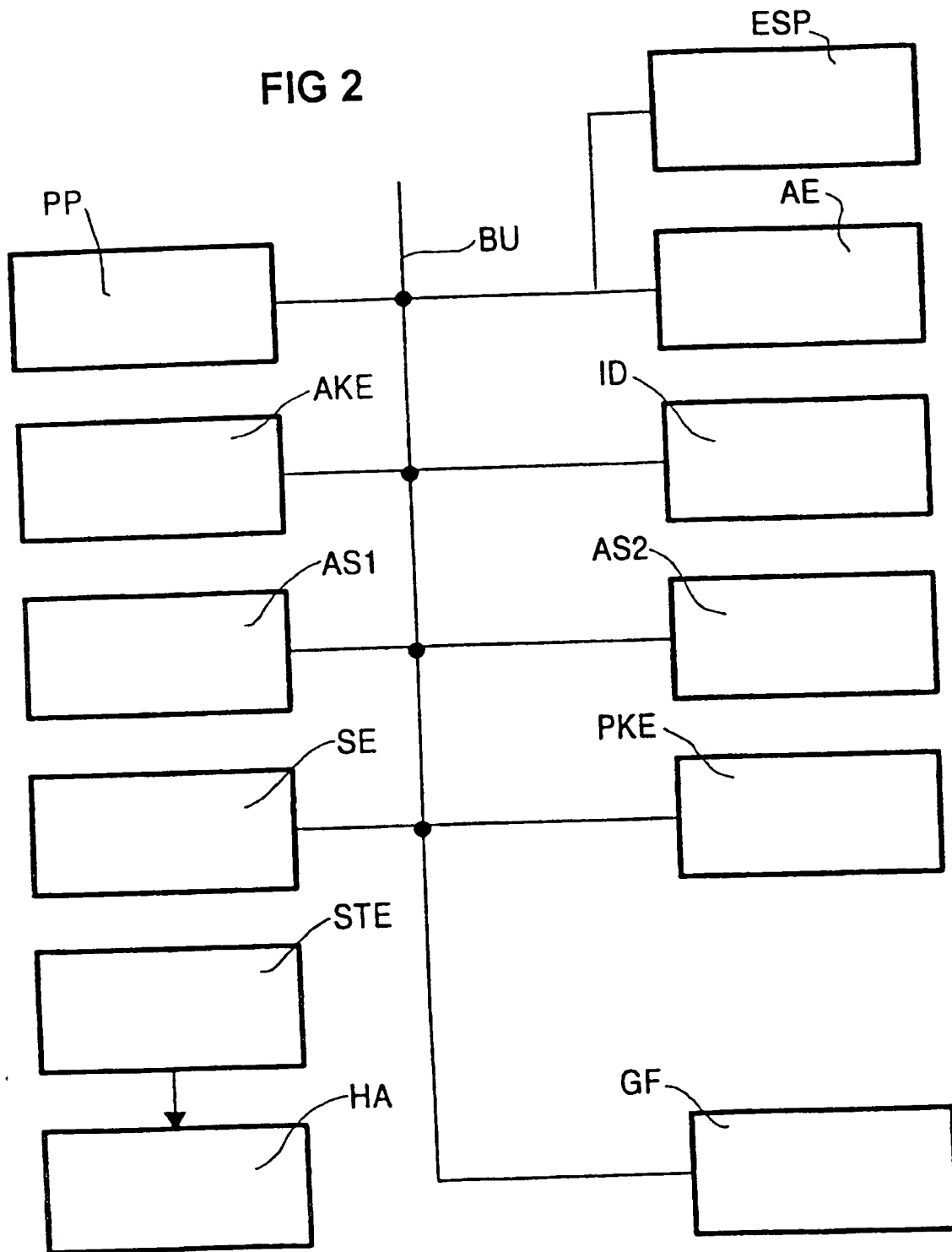
10. Verfahren nach Anspruch 8 oder 9,  
bei dem nach der Spracherkennung eine Nachbearbeitung des erkannten Sprachsignals erfolgt.

FIG 1



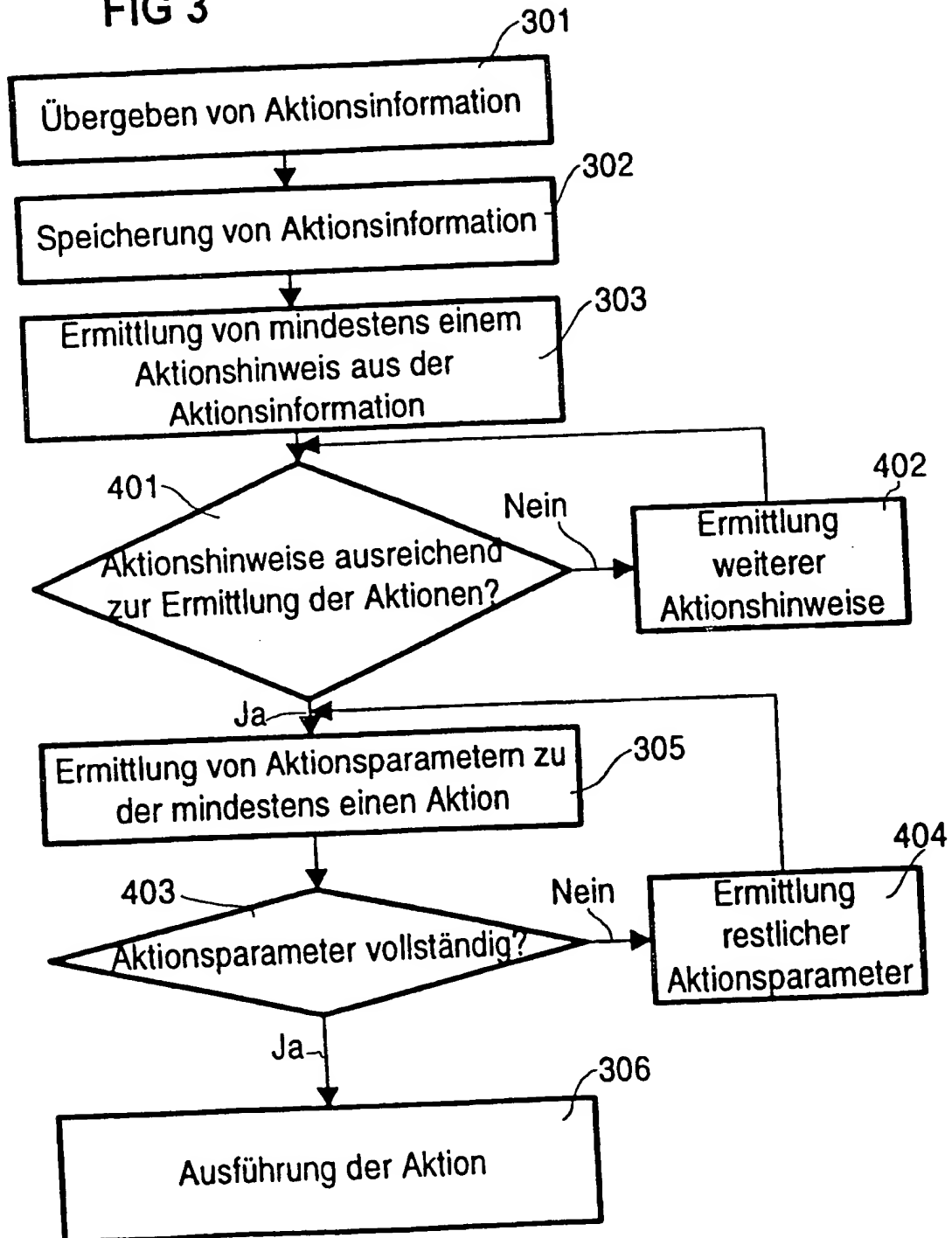
2 / 4

FIG 2



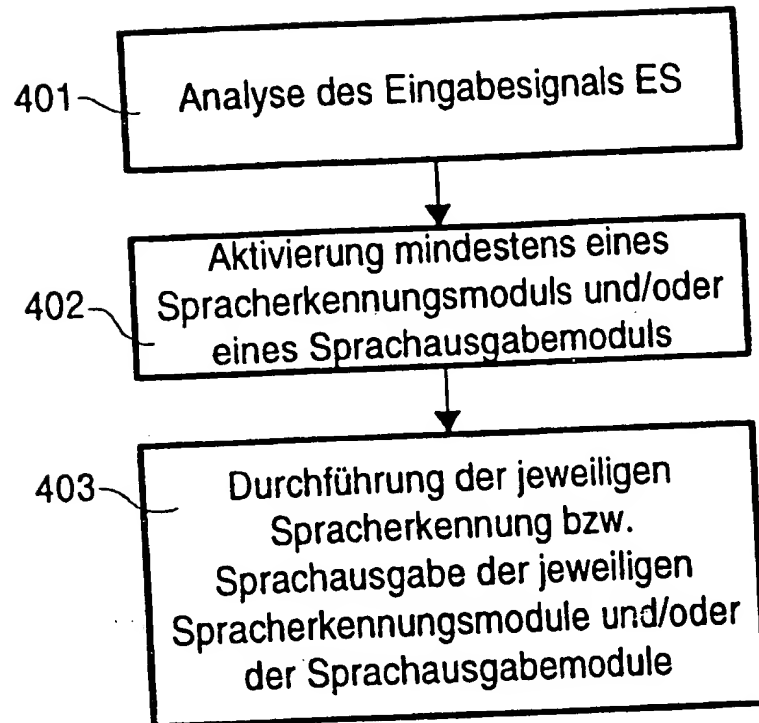
3 / 4

FIG 3



4 / 4

FIG 4



# INTERNATIONAL SEARCH REPORT

International Application No  
PCT/DE 97/01920

A. CLASSIFICATION OF SUBJECT MATTER  
IPC 6 G10L3/00

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)  
IPC 6 G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, X	EP 0 730 261 A (SEIKO EPSON CORP) 4 September 1996 see abstract see claims 1-4; figures 1-6 see column 15, line 47 - column 17, line 5 ---	1, 2, 5, 8
X	WO 96 18260 A (UNIV OXFORD BROOKES ; BROWNSEY KENNETH (GB); ZAJICEK MARY (GB)) 13 June 1996 see abstract; claims 1, 2, 5-7 ---	1, 2, 5, 8
A	EP 0 607 615 A (TOKYO SHIBAURA ELECTRIC CO ; TOSHIBA SOFT ENGINEERING COMPA (JP)) 27 July 1994 see abstract; claims 1-3, 28-30, 57-60; figures 56, 57, 102 --- -/--	1, 2, 5, 8

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

### \* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

5 January 1998

Date of mailing of the international search report

04/02/1998

Name and mailing address of the ISA  
European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Van Doremalen, J

# INTERNATIONAL SEARCH REPORT

Inter national Application No  
PCT/DE 97/01920

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	DE 44 12 309 A (TELEFONBAU & NORMALZEIT GMBH) 12 October 1995 cited in the application see abstract; claims 1-4 ---	1-9
A	GB 2 119 208 A (GEN ELECTRIC CO PLC) 9 November 1983 cited in the application see abstract; claims 1,4,5,9 ---	1,8
A	DATABASE INSPEC INSTITUTE OF ELECTRICAL ENGINEERS, STEVENAGE, GB Inspec No. 2578705, BERGERON L E ET AL: "Voice applications for the office using digital signal processing" XP002051225 see abstract & CONFERENCE RECORD EIGHTEENTH ASILOMAR CONFERENCE ON CIRCUITS, SYSTEMS AND COMPUTERS (CAT. NO. 85CH2200-4), PACIFIC GROVE, CA, USA, 5-7 NOV. 1984, ISBN 0-8186-0673-8, 1985, WASHINGTON, DC, USA, IEEE COMPUT. SOC. PRESS, USA, pages 238-242, -----	1,8



# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/DE 97/01920

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0730261 A	04-09-96	JP 8297498 A CN 1142647 A	12-11-96 12-02-97
WO 9618260 A	13-06-96	AU 4182596 A	26-06-96
EP 0607615 A	27-07-94	JP 7140998 A US 5632002 A	02-06-95 20-05-97
DE 4412309 A	12-10-95	NONE	
GB 2119208 A	09-11-83	CA 1201809 A EP 0093022 A	11-03-86 02-11-83

# INTERNATIONALER RECHERCHENBERICHT

Internationales Aktenzeichen  
PCT/DE 97/01920

A. KLASSIFIZIERUNG DES ANMELDUNGSGEGENSTANDES  
IPK 6 G10L3/00

Nach der internationalen Patentklassifikation (IPK) oder nach der nationalen Klassifikation und der IPK

## B. RECHERCHIERTE GEBIETE

Recherchierter Mindestprüfstoff (Klassifikationssystem und Klassifikationssymbole)  
IPK 6 G10L

Recherchierte aber nicht zum Mindestprüfstoff gehörende Veröffentlichungen, soweit diese unter die recherchierten Gebiete fallen

Während der internationalen Recherche konsultierte elektronische Datenbank (Name der Datenbank und evtl. verwendete Suchbegriffe)

## C. ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
-----------	--	--------------------

P, X	EP 0 730 261 A (SEIKO EPSON CORP) 4. September 1996 siehe Zusammenfassung siehe Ansprüche 1-4; Abbildungen 1-6 siehe Spalte 15, Zeile 47 - Spalte 17, Zeile 5	1, 2, 5, 8
X	WO 96 18260 A (UNIV OXFORD BROOKES ; BROWNSEY KENNETH (GB); ZAJICEK MARY (GB)) 13. Juni 1996 siehe Zusammenfassung; Ansprüche 1, 2, 5-7	1, 2, 5, 8
A	EP 0 607 615 A (TOKYO SHIBAURA ELECTRIC CO ; TOSHIBA SOFT ENGINEERING COMPA (JP)) 27. Juli 1994 siehe Zusammenfassung; Ansprüche 1-3, 28-30, 57-60; Abbildungen 56, 57, 102	1, 2, 5, 8

-/-

☒ Weitere Veröffentlichungen sind der Fortsetzung von Feld C zu entnehmen

☒ Siehe Anhang Patentfamilie

\* Besondere Kategorien von angegebenen Veröffentlichungen :

"A" Veröffentlichung, die den allgemeinen Stand der Technik definiert, aber nicht als besonders bedeutsam anzusehen ist

"E" älteres Dokument, das jedoch erst am oder nach dem internationalen Anmeldedatum veröffentlicht worden ist

"L" Veröffentlichung, die geeignet ist, einen Prioritätsanspruch zweifelhaft erscheinen zu lassen, oder durch die das Veröffentlichungsdatum einer anderen im Recherchenbericht genannten Veröffentlichung belegt werden soll oder die aus einem anderen besonderen Grund angegeben ist (wie ausgeführt)

"O" Veröffentlichung, die sich auf eine mündliche Offenbarung, eine Benutzung, eine Ausstellung oder andere Maßnahmen bezieht

"P" Veröffentlichung, die vor dem internationalen Anmeldedatum, aber nach dem beanspruchten Prioritätsdatum veröffentlicht worden ist

"T" Spätere Veröffentlichung, die nach dem internationalen Anmeldedatum oder dem Prioritätsdatum veröffentlicht worden ist und mit der Anmeldung nicht kollidiert, sondern nur zum Verständnis des der Erfindung zugrundeliegenden Prinzips oder der ihr zugrundeliegenden Theorie angegeben ist

"X" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann allein aufgrund dieser Veröffentlichung nicht als neu oder auf erfinderischer Tätigkeit beruhend betrachtet werden

"Y" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann nicht als auf erfinderischer Tätigkeit beruhend betrachtet werden, wenn die Veröffentlichung mit einer oder mehreren anderen Veröffentlichungen dieser Kategorie in Verbindung gebracht wird und diese Verbindung für einen Fachmann naheliegend ist

"G" Veröffentlichung, die Mitglied derselben Patentfamilie ist

Datum des Abschlusses der internationalen Recherche

5. Januar 1998

Absendedatum des internationalen Recherchenberichts

04/02/1998

Name und Postanschrift der Internationalen Recherchenbehörde  
Europäisches Patentamt, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Bevollmächtigter Bediensteter

Van Doremalen, J

# INTERNATIONALER RECHERCHENBERICHT

Inter. ...ionales Aktenzeichen  
PCT/DE 97/01920

## C.(Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
A	DE 44 12 309 A (TELEFONBAU & NORMALZEIT GMBH) 12.Oktober 1995 in der Anmeldung erwähnt siehe Zusammenfassung; Ansprüche 1-4 ---	1-9
A	GB 2 119 208 A (GEN ELECTRIC CO PLC) 9.November 1983 in der Anmeldung erwähnt siehe Zusammenfassung; Ansprüche 1,4,5,9 ---	1,8
A	DATABASE INSPEC INSTITUTE OF ELECTRICAL ENGINEERS, STEVENAGE, GB Inspec No. 2578705, BERGERON L E ET AL: "Voice applications for the office using digital signal processing" XP002051225 siehe Zusammenfassung & CONFERENCE RECORD EIGHTEENTH ASILOMAR CONFERENCE ON CIRCUITS, SYSTEMS AND COMPUTERS (CAT. NO. 85CH2200-4), PACIFIC GROVE, CA, USA, 5-7 NOV. 1984, ISBN 0-8186-0673-8, 1985, WASHINGTON, DC, USA, IEEE COMPUT. SOC. PRESS, USA, Seiten 238-242, -----	1,8

# INTERNATIONALER RECHERCHENBERICHT

Angaben zu Veröffentlichungen, die zur selben Patentfamilie gehören

Internationales Aktenzeichen  
PCT/DE 97/01920

Im Recherchenbericht angeführtes Patentdokument	Datum der Veröffentlichung	Mitglied(er) der Patentfamilie	Datum der Veröffentlichung
EP 0730261 A	04-09-96	JP 8297498 A CN 1142647 A	12-11-96 12-02-97
WO 9618260 A	13-06-96	AU 4182596 A	26-06-96
EP 0607615 A	27-07-94	JP 7140998 A US 5632002 A	02-06-95 20-05-97
DE 4412309 A	12-10-95	KEINE	
GB 2119208 A	09-11-83	CA 1201809 A EP 0093022 A	11-03-86 02-11-83